

PERMANENT GENETIC RESOURCES NOTE

Characterization of 22 novel single nucleotide polymorphism markers in steelhead and rainbow trout

NATHAN R. CAMPBELL,* KEN OVERTURF† and SHAWN R. NARUM*

**Columbia River Intertribal Fish Commission, 3059 F National Fish Hatchery Rd, Hagerman, ID 83332, USA, †US Department of Agriculture, 3059 F National Fish Hatchery Rd, Hagerman, ID 83332, USA*

Abstract

Thirty-two individuals representing coastal and inland populations of steelhead and rainbow trout (*Oncorhynchus mykiss*) were sequenced at 18 expressed sequence tags and nine microsatellite loci to identify single nucleotide polymorphisms. A total of 98 polymorphisms were discovered during the screen and 22 were developed into 5' exonuclease assays (Taqman assays). Genotypes from TaqMan assays were compared to sequence data from individuals in the ascertainment panel to confirm proper allele designations. A larger number of samples ($n = 192$) from six regions were tested with the validated assays. Per-locus F_{ST} values ranged from 0.001 to 0.414.

Keywords: 5' exonuclease assay, *Oncorhynchus mykiss*, rainbow trout, SNP, steelhead, Taqman

Received 1 August 2008; revision accepted 22 August 2008

Steelhead and rainbow trout are anadromous and resident life-history types of *Oncorhynchus mykiss* that are native to the west coast of North America, but are now common to many cold-water lakes and streams the world over due to human introduction. This species is commonly reared in hatchery settings and have become a major contributor to the world's growing aquaculture industry, with commercial farms in 64 countries on all continents except Antarctica (United Nations Food and Agriculture; <http://www.fao.org>). However, despite the species' widespread abundance, several native populations of *O. mykiss* have been listed as threatened or endangered by the US Fish and Wildlife Service (<http://www.fws.gov/endangered/wildlife.html>).

Genetic markers such as microsatellites have been used in recent years in both population studies (e.g. Narum *et al.* 2004) and as broodstock markers in the aquaculture industry (Silverstein *et al.* 2004). Technical issues with using microsatellite markers (such as intensive interlaboratory standardization efforts) have led many researchers to explore the use of single nucleotide polymorphisms (SNPs) as an alternative genetic marker (Morin *et al.* 2004). Despite having only two possible alleles per locus, SNPs offer a potential cost savings over microsatellite markers as well as easily transferable data between laboratories. Here we describe the identification of 98 novel SNP sites and the development of 22 genotyping assays based on the 5' nuclease reaction.

Eighteen expressed sequence tags and nine microsatellite loci from *O. mykiss* were selected from GenBank and TIGR databases for primer design. Primers were designed using the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to generate a product of about 400 bp. Since primer design was based on mRNA sequences, target sizes of 400 bp allowed for flexibility in actual amplicon size in the event that an intron region was amplified in genomic DNA. Each set of primers was then tested by polymerase chain reaction (PCR) on four *O. mykiss* genomic DNA extracts under the following conditions: 1× Thermophilic DNA polymerase buffer (Promega), 2 mM $MgCl_2$, 0.25 mM dNTPs, 0.1 mg/mL BSA, and 1 U/rxn AmpliTaq polymerase (Applied Biosystems) with 2 µL genomic DNA (extracted using QIAGEN DNeasy 96 kits) in 12 µL total volume. Thermal cycler conditions for PCR were typically 94 °C for 1 min, 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min/kb, with steps 2–4 repeated for 35 cycles followed by a cool-down to 4 °C (performed on MJ Research thermal cyclers). Amplified fragments were evaluated by agarose gel electrophoresis for clean amplification of a single product. Primer pairs that produced clean fragments were selected for amplification and sequencing of an ascertainment panel of 32 individuals with diverse genetic backgrounds. Fifteen loci were successfully amplified and sequenced using BigDye version 1.1 chemistry and a 3730 DNA sequencer (Applied Biosystems). Successful PCR primer sequences and accession numbers are listed in Table S1, Supporting information. Following electrophoresis,

Correspondence: Shawn R. Narum, Fax: 208-837-6047; E-mail: nars@critfc.org

chromatogram data were analysed by first using the Sequencing Analysis program (ABI) to assign base calls. Then, Sequencher version 4.6 (Gene Codes) was used to align and edit the data. Sequences containing introns were annotated and submitted to GenBank (Accession nos EF675993, EF675994, EF675995, EF675996, EU682503, and EU682504). Consensus sequences from all loci were submitted to dbSTS and all observed sequence variations were submitted to dbSNP (Table S1). BLAST search queries (<http://www.ncbi.nlm.nih.gov/BLAST/>) of consensus sequences confirmed amplification of all target loci. The screen revealed 89 SNP sites and two microsatellite repeats linked to known genes (Table S1). Nine SNP sites were discovered in microsatellite flanking regions. Twenty-three SNPs were chosen based on minor allele frequency data for development of 5' exonuclease assays (Taqman assays).

Custom TaqMan allelic discrimination assays (5' exonuclease assays) were ordered through Applied Biosystems. Primer and probe mix (40×) for each submitted SNP was designed and produced by Applied Biosystems. Genotyping

assays were performed in 384-well plates using 1 µL evaporated *O. mykiss* genomic DNA and 5 µL 1× TaqMan Universal PCR Master Mix (No AmpErase UNG, ABI) with 1× primer-probe mixture. Amplification was performed using MJ Research and Applied Biosystems thermal cyclers with 384-well blocks using standard two-step cycling for 50 cycles. End point reads of each reaction were collected using an Applied Biosystems 9700HT instrument and data were analysed using the allelic discrimination function of Applied Biosystems SDS version 2.1 software. All assays were validated by comparing allele calls to sequencing data for agreement. Twenty-two assays showed complete agreement with sequencing data and one was rejected due to poor performance. A list of primers and probes for each validated SNP assay is shown in Table 1. Assay annealing temperature was increased to 62 °C for assay *Omy_aspAT-123* and extension time was increased for assay *Omy_ots208-138* to 2 min to reduce genotyping error.

Validated assays were used to genotype a set of 96 *O. mykiss* samples which included representatives from the

Table 1 Twenty-two validated SNP assays in *Oncorhynchus mykiss*. Heterozygosity and F_{ST} estimates exclude Columbia mixture samples collected at Bonneville Dam

Locus (locus name, assay name, variation, reference GenBank nos)	Oligos (forward primer, reverse primer, allele-specific probe 1, allele specific probe 2)	T_a (assay annealing temp. °C)	N	Heterozygosity		
				H_E	H_O	F_{ST}
Cyclooxygenase-1 <i>Omy_cox1-221</i> (T/A) AJ299018	CACTGAACGTGAAGCCATTGTGATT GCAACATGGGAATGATTTCATAAATGCA VIC-CGGTAAGACCATTAAAA FAM-CGGTAAGACCATTAAAA	60	94	0.439	0.404	0.119
Myostatin 1a <i>Omy_myo1a-264</i> (A/C) AF273035	CCCATCAACATGCTCTACTTCAAC CGGTCCACCACCATGGA VIC-AAGAGCAGATAATCTAC FAM-AGAGCAGATCATCTAC	60	94	0.335	0.383	0.158
Myostatin 1b <i>Omy_myo1b-111</i> (G/C) AF273036	ACCTGGTGAACAAGGCTAACC TGGATGGGATCTTGCCGTAGA VIC-ATGTCCCGATCAAC FAM-ATGTCCCGCATCAAC	60	93	0.124	0.075	0.126
nkef <i>Omy_nkef-241</i> (C/A) AF250195	AGTGTCAITGATGTCGGCCTATTTT AAACGAATGTCCACCTCAGATGTT VIC-CTTCTGTATCATTTTTG FAM-TCTTCTGTATAATTTTTG	60	93	0.333	0.312	0.279
nkef <i>Omy_nkef-308</i> (T/G) AF250195	CTTTAGGCCACAGATGGGATAGTC GCCACATGTTTTGTCTGTAAAGC VIC-TTCTGTTCTTAATTICA FAM-TTCTGTTCTCAATTICA	60	94	0.246	0.309	0.226
myo D <i>Omy_myoD-178</i> (A/C) Z46924	TGGCAAAGCTGTCAATTCCTTCTAAT GGTCAAATATTTCAATTACGATTACACTTAGGC VIC-TTTTATGAGATATAATTTCC FAM-TTTTATGAGATATCATTTCC	60	94	0.111	0.128	0.032
nramp-alpha <i>Omy_nramp-146</i> (G/A) EF675993	TGAGAGTGCACATTGTATTGTTAACCTTT CACATCCCTACTGACAAAACACTGA VIC-CGTGTGTTGGTGTTTT FAM-CGTGTGTTGATGTTTT	60	94	0.110	0.106	0.414
Ots-474 <i>Omy_Ots474-67</i> (G/A) AF393200	AGCTCTGGACATTTTATCACACACAA GGAGCTTGCTAGTCTAAACAGAT VIC-AAGAACCGACTGATCC FAM-TTATAAGAACCAACTGATCC	60	95	0.027	0.032	0.009

Table 1 Continued

Locus (locus name, assay name, variation, reference GenBank nos)	Oligos (forward primer, reverse primer, allele-specific probe 1, allele specific probe 2)	T_a (assay annealing temp. °C)	N	Heterozygosity		F_{ST}
				H_E	H_O	
Ots-249 <i>Omy</i> _Ots249-227 (C/T) AF393192	CCCCTAGATTAAACCTGTCCAGTCT CTATCTATCTATCTATCTATCTATC TATCTATCTATCTACTTACTGAGA VIC-CCCTCTGAGAACTAC FAM-CCTCTGAAAACTAC	60	94	0.436	0.532	0.131
Ots-208 <i>Omy</i> _Ots208-138 (G/T) AF393187	ACTGCAGCTTGGTCCATTGATAAT ACATACTTCAACTTCCCTTGAACCC VIC-TTAACATCAAAGTAACATACCAA FAM-ACATCAAAGTAACATAACAA GTGCCAAACTCTTCTATCCTCTCT	60	91	0.420	0.308	0.030
Ogo-4 <i>Omy</i> _Ogo4-164 (G/T) AF009796	TTGGTAGAATAATTAGTGTGT GTTAAATAAGCATTTG VIC-AAGGTTTATGCAGGTTAT FAM-AAAGGTTTATGAAGGTTAT	58	94	0.031	0.032	0.182
Ogo-4 <i>Omy</i> _Ogo4-212 (T/C) AF009796	TCCTCTCTCCCATTCATCACTAATGA AGACAGTAACAAGCCCTCAAACCTGA VIC-CATTTGATGAGACATCTT FAM-ATTTGATGAGGCATCTT	60	95	0.371	0.453	0.181
Aldolase B <i>Omy</i> _aldB-165 (C/G) EF675994	GCGTTAGGTGGATTTGAAGGAGTAA AGGAAGGTGATGCCTGAGAGA VIC-ATGCTAAAATGAATCCCCACCA FAM-CTAAAAATGAATCGCCACCA	60	94	0.373	0.383	0.216
Aldolase B <i>Omy</i> _aldB-414 (A/T) EF675994	TCCAGAGCCAAGGTAAGATGTT CACCTGTAAGTAAGTTGTCCACTCT VIC-CCCTAGAAATAGCCCC FAM-CCTAGAAAAGCCCC	60	94	0.096	0.096	0.086
Aspartate aminotransferase <i>Omy</i> _aspAT-123 (T/C) EF675996	GTTTGCCCATTTCACTGATGCT AGGAGACCACTCCAAAGAGAACT VIC-CCTTCCTAGGCAGTCAG FAM-TTCCTGGGCAGTCAG	62	95	0.178	0.221	0.002
Aspartate aminotransferase <i>Omy</i> _aspAT-413 (G/T) EF675996	CCTGAACAGGTACACACAAACGA CCAACTGATGAATATGACCAACAT TAATATAATAAATGTC VIC-CACTCTTTATATCCACACCTGT FAM-CTCTTTATATCCAAACCTGT	60	93	0.036	0.038	0.028
Glutamate dehydrogenase <i>Omy</i> _gdh-271 (C/T) AJ556997	AGGTCACTCTACTTACAGTATAAAGCAGT GTCATGTC AACAGAGTAACATAATAAATCTGC VIC-TCACCCCTGAAGGTAGAC FAM-TCACCCCTGAAATGTAGAC	60	95	0.130	0.126	0.001
NaK ATPase a3 <i>Omy</i> _NaK-a3-50 (A/C) AY319388	GTTGAGCGTGTATGGGAAAAGAG TTGCATCGGCTTTCTGAAAACC VIC-CACTCTGTTTCTTTCTTT FAM-TCTGTTTCCGTTCTTT	60	96	0.214	0.203	0.004
Growth hormone 1 <i>Omy</i> _gh-334 (A/*) EU682503	CCAAATGAGAAGTCACATCAATGCA CTTTGTGTAGCATAAAATGAATCAATCACTCA VIC-CAGGTAGATTTTTTTAAATG FAM-CAGGTAGATTTTTTTAAATG	62	95	0.036	0.026	0.112
Growth hormone 1 <i>Omy</i> _gh-475 (C/T) EU682503	AAGTTACCAGAATTTTGCAAACCTCAACT CCATATTTTGAGGTGTAGCTTTACCCCT VIC-CTGAAACTCATGGTATACA FAM-CTGAAACTCATGATATACA	60	95	0.130	0.100	0.230
Growth hormone 1 <i>Omy</i> _gh-1093 (C/T) EU682503	GTGACACCCCAITCAATGACTGA GCACACGTAGCAAAAGACACATG VIC-CATAAATGTCCTTGAATGG FAM-CATAAATGTCCTTAAATGG	60	96	0.180	0.130	0.033
Acidic ribosomal phosphoprotein <i>Omy</i> _arp-630 (G/A) EU682504	CTGCACAACTTGTTCCTGCTATT ACCAAGTGTCCCTGTAAGCC VIC-CCGCTCCGTCTGCT FAM-CCGCTCTGTCTGCT	60	93	0.233	0.183	0.048

Table 2 Observed allele frequencies of six collections of *Oncorhynchus mykiss* for the minor allele at 22 SNP loci

Locus	Population							Overall
	Upper Columbia	Middle Columbia	Snake River	Lower Columbia	SE Alaska	McCloud strain	Columbia mixture	
N	16	24	16	16	16	8	96	192
<i>Omy_cox1-221</i>	0.219	0.375	0.437	0.679	0.688	0.250	0.366	0.406
<i>Omy_myo1a-264</i>	0.156	0.125	0.200	0.233	0.219	0.312	0.152	0.220
<i>Omy_myo1b-111</i>	0.000	0.000	0.000	0.100	0.062	0.286	0.011	0.029
<i>Omy_nkef-241</i>	0.474	0.370	0.600	0.133	0.844	0.000	0.510	0.473
<i>Omy_nkef-308</i>	0.437	0.542	0.400	0.300	0.094	1.000	0.318	0.371
<i>Omy_myoD-178</i>	0.062	0.062	0.179	0.062	0.000	0.000	0.141	0.103
<i>Omy_nramp-146</i>	0.031	0.000	0.031	0.000	0.000	0.500	0.010	0.032
<i>Omy_Ots474-67</i>	0.062	0.021	0.000	0.000	0.000	0.000	0.010	0.013
<i>Omy_Ots249-227</i>	0.344	0.458	0.312	0.833	0.467	0.250	0.321	0.389
<i>Omy_Ots208-138</i>	0.200	0.370	0.267	0.533	0.219	0.286	0.333	0.326
<i>Omy_Ogo4-212</i>	0.594	0.354	0.469	0.467	1.000	0.125	0.374	0.363
<i>Omy_Ogo4-164</i>	0.031	0.021	0.000	0.000	0.000	0.286	0.011	0.021
<i>Omy_aldB-165</i>	0.375	0.437	0.233	0.733	0.781	0.937	0.277	0.410
<i>Omy_aldB-414</i>	0.125	0.187	0.233	0.000	0.000	0.000	0.182	0.145
<i>Omy_aspAT-123</i>	0.125	0.125	0.187	0.067	0.094	0.000	0.302	0.207
<i>Omy_aspAT-413</i>	0.031	0.022	0.067	0.000	0.125	0.000	0.130	0.085
<i>Omy_gdh-271</i>	0.062	0.042	0.062	0.156	0.100	0.000	0.089	0.081
<i>Omy_gh-334</i>	0.000	0.000	0.167	0.000	0.000	0.125	0.005	0.018
<i>Omy_gh-475</i>	0.062	0.125	0.100	0.167	0.031	0.687	0.124	0.142
<i>Omy_gh-1093</i>	0.312	0.354	0.187	0.219	0.156	0.000	0.297	0.266
<i>Omy_arp-630</i>	0.344	0.341	0.467	0.562	0.156	0.312	0.400	0.384
<i>Omy_NaKATPa3-50</i>	0.375	0.312	0.312	0.143	0.406	0.250	0.344	0.325

upper, middle, and lower Columbia River as well as the Snake River, southeast Alaska, and Goldendale Hatchery (McCloud aquaculture strain). A second set of 96 anadromous *O. mykiss* samples collected at Bonneville Dam on the Columbia River was also genotyped (labelled Columbia River mixture in Table 2). Microsatellite Analyser (Deiringer & Schlötterer 2003) was used to calculate allele frequencies for both sets of samples as well as expected and observed heterozygosities and Global F_{ST} values for each locus (Tables 1 and 2). Since samples collected at Bonneville Dam were an admixed collection of upstream populations, samples were not included in F_{ST} or heterozygosity estimates.

The observed heterozygosity of loci ranged from 0.026 to 0.532 across regions (Table 1). Per-locus estimates of F_{ST} ranged from 0.001 (*Omy_gdh-271*) to 0.414 (*Omy_nramp-146*; Table 1) and averaged 0.120. Frequency of the minor allele in some SNPs was highly variable among regions genotyped (Table 2). For example, *Omy_nramp-146* had minor allele frequencies ranging from 0.00 to 0.50. This genetic variation indicates that these novel SNP markers may be useful for studies in aquaculture and conservation genetics of *O. mykiss*. However, potential for ascertainment bias must be considered if these markers are applied to populations other than those included in the sequencing panel (Luikart *et al.* 2003).

Acknowledgements

Special thanks to Ruth Philips and Jennifer DeKoning for their assistance in our experimental approach to SNP discovery.

References

- Deiringer D, Schlötterer C (2003) Microsatellite Analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews*, **4**, 981–994.
- Morin PA, Luikart G, Wayne RK, SNP Workshop Group (2004) Application of SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.
- Narum SR, Contor C, Talbot A, Powell MS (2004) Genetic divergence of sympatric resident and anadromous forms of *Oncorhynchus mykiss* in the Walla Walla river, U.S.A. *Journal of Fish Biology*, **65**, 471–488.
- Silverstein JT, Rexroad CE, King TL (2004) Genetic variation measured by microsatellites among three strains of domesticated rainbow trout (*Oncorhynchus mykiss*, Walbaum). *Aquaculture Research*, **35**, 40–48.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Table S1 PCR/sequencing primers for all loci screened and database references for sequences used to generate them are shown along with all observed variations and their allele frequencies. dbSTS and dbSNP reference numbers are also shown. Minor allele frequencies were not calculated for observed microsatellite repeats (denoted with a 'μ'). Small deletions are denoted with a 'd' and larger indels are denoted with an 'i'.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.